Assignment 5: Data Visualization

Aubrey Knier

OVERVIEW

This exercise accompanies the lessons in Environmental Data Analytics on Data Visualization

Directions

1. Change "Student Name" on line 3 (above) with your name.
2. Work through the steps, **creating code and output** that fulfill each instruction.
3. Be sure to **answer the questions** in this assignment document.
4. When you have completed the assignment, **Knit** the text and code into a single PDF file.
5. After Knitting, submit the completed exercise (PDF file) to the dropbox in Sakai. Add your last name into the file name (e.g., "Fay_A05_DataVisualization.Rmd") prior to submission.

The completed exercise is due on Monday, February 14 at 7:00 pm.

Set up your session

1. Set up your session. Verify your working directory and load the tidyverse and cowplot packages. Upload the NTL-LTER processed data files for nutrients and chemistry/physics for Peter and Paul Lakes (use the tidy [`NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Processed.csv`] version) and the processed data file for the Niwot Ridge litter dataset (use the [`NEON_NIWO_Litter_mass_trap_Processed.csv`] version).

2. Make sure R is reading dates as date format; if not change the format to date.

```
#1
setwd("/Users/aubreyknier/Desktop/Spring 2022/ENV872_EDA/Environmental_Data_Analytics_2022")
getwd()
```

```
## [1] "/Users/aubreyknier/Desktop/Spring 2022/ENV872_EDA/Environmental_Data_Analytics_2022"
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(cowplot)

PeterPaul.chem.nutrients.data <-
  read.csv("./Data/Processed/NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Processed.csv")

NIWO.litter.data <-
  read.csv("./Data/Processed/NEON_NIWO_Litter_mass_trap_Processed.csv")

#2

#Peter Paul Lakes dates
class(PeterPaul.chem.nutrients.data$sampledate)
```

```
## [1] "character"
```

```
PeterPaul.chem.nutrients.data$sampledate <- as.Date(
  PeterPaul.chem.nutrients.data$sampledate, format = "%Y-%m-%d")
class(PeterPaul.chem.nutrients.data$sampledate)
```

```
## [1] "Date"
```

```
#NIWOT ridge dates
class(NIWO.litter.data$collectDate)
```

```
## [1] "character"
```

```
NIWO.litter.data$collectDate <- as.Date(
  NIWO.litter.data$collectDate, format = "%Y-%m-%d")
class(NIWO.litter.data$collectDate)
```

```
## [1] "Date"
```

Define your theme

3. Build a theme and set it as your default theme.

```
#3
mytheme <- theme_linedraw(base_size = 14) +
  theme(axis.text = element_text(color = "grey15", face="bold"),
        legend.position = "top")

theme_set(mytheme)
```

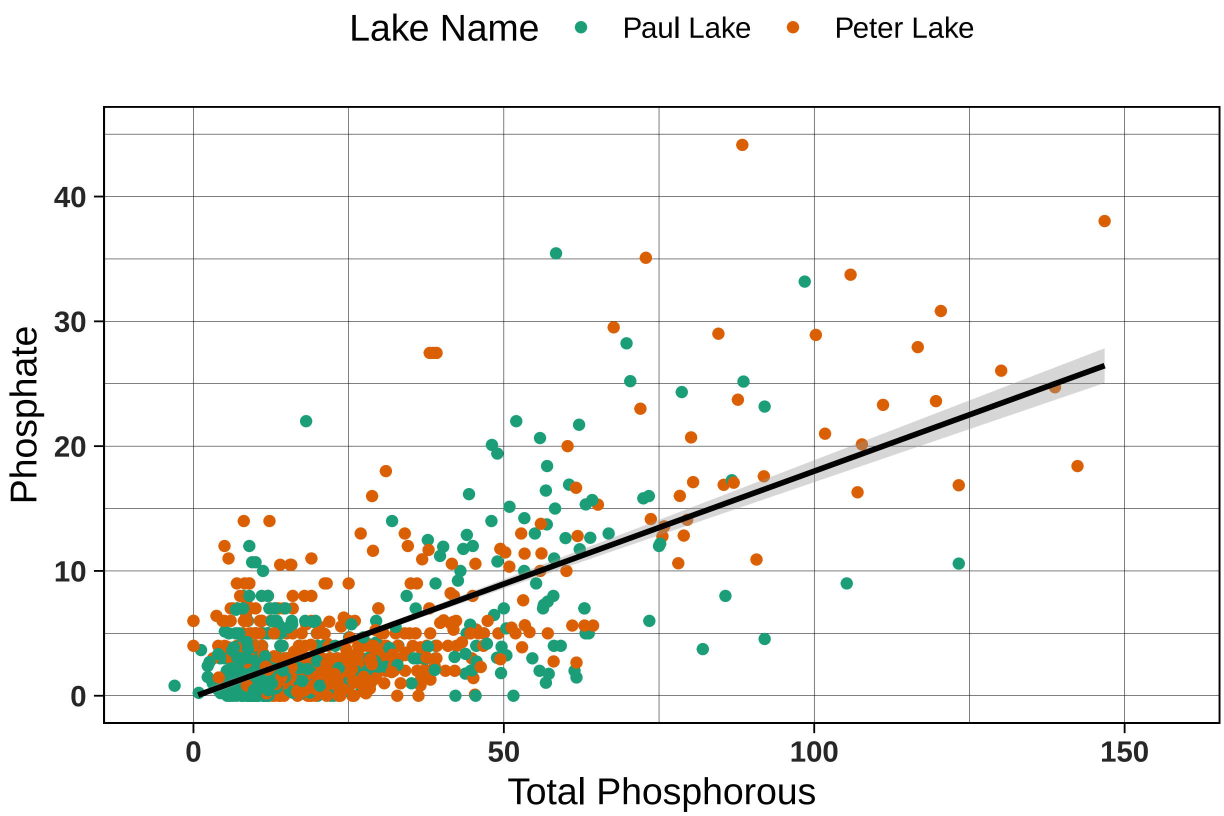
Create graphs

For numbers 4-7, create ggplot graphs and adjust aesthetics to follow best practices for data visualization. Ensure your theme, color palettes, axes, and additional aesthetics are edited accordingly.

4. [NTL-LTER] Plot total phosphorus (`tp_ug`) by phosphate (`po4`), with separate aesthetics for Peter and Paul lakes. Add a line of best fit and color it black. Adjust your axes to hide extreme values (hint: change the limits using `xlim()` and `ylim()`).

```
#4
library(ggplot2)
library(RColorBrewer)
#display.brewer.all()



NTL_LTER_phos_plot <-
ggplot(PeterPaul.chem.nutrients.data, aes(x=tp_ug, y=po4, color=lakename)) +
  geom_point() + ylim(0,45) +
    xlab("Total Phosphorous") + ylab("Phosphate") +
      labs(color="Lake Name") +
        geom_smooth(method="lm", color="black") +
         scale_color_brewer(palette = "Dark2")
print(NTL_LTER_phos_plot)
```

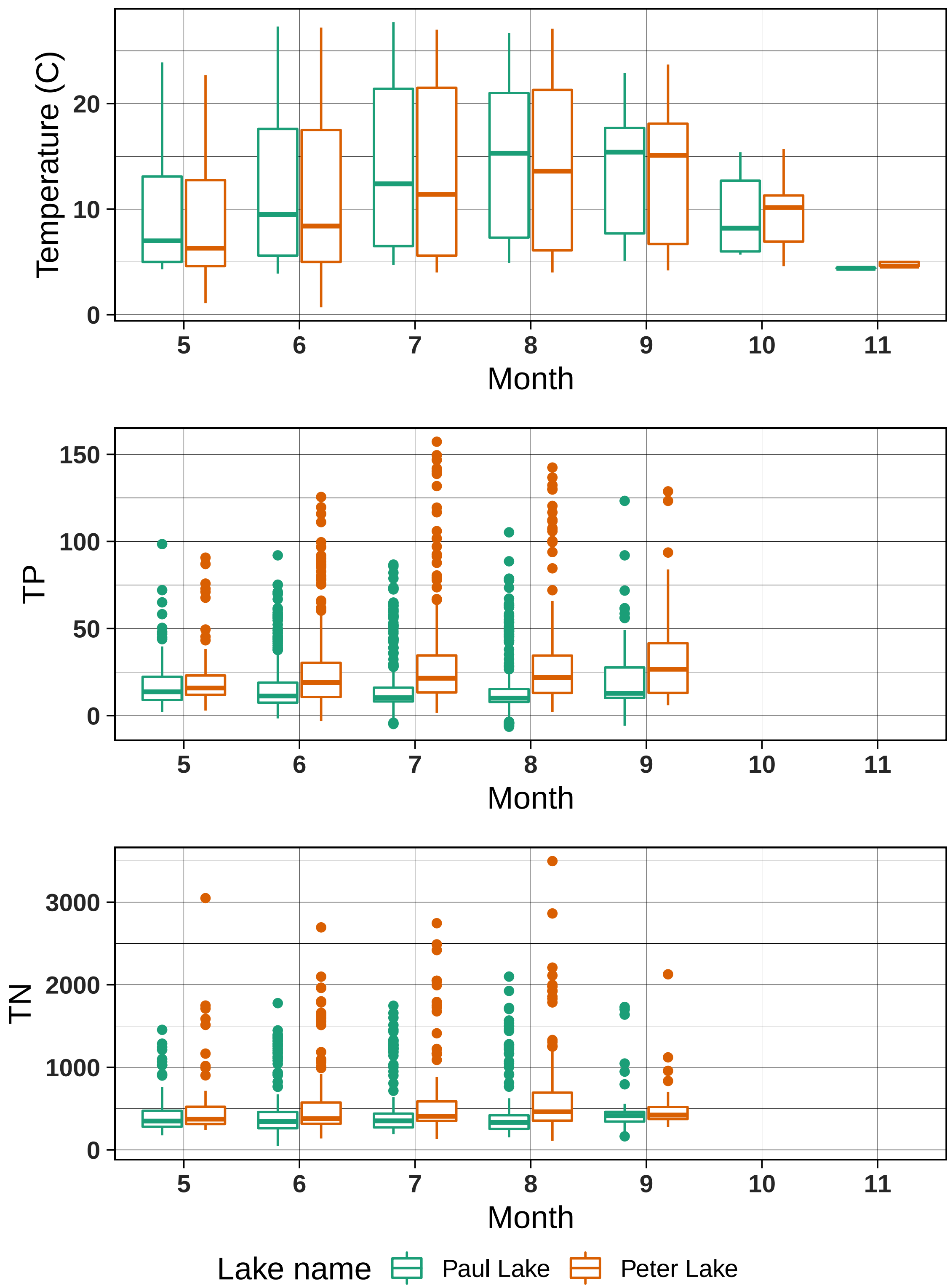
```
## `geom_smooth()` using formula 'y ~ x'

## Warning: Removed 21947 rows containing non-finite values (stat_smooth).

## Warning: Removed 21947 rows containing missing values (geom_point).

## Warning: Removed 2 rows containing missing values (geom_smooth).
```



5. [NTL-LTER] Make three separate boxplots of (a) temperature, (b) TP, and (c) TN, with month as the x axis and lake as a color aesthetic. Then, create a cowplot that combines the three graphs. Make sure that only one legend is present and that graph axes are aligned.

```
#5
#display.brewer.all()


PeterPaul.chem.nutrients.data$month <- as.factor(PeterPaul.chem.nutrients.data$month)

#a. temperature
NTL_LTER_temp_boxplot <-
ggplot(PeterPaul.chem.nutrients.data, aes(x=month, y=temperature_C)) +
  geom_boxplot(aes(color=lakename)) +
    ylab("Temperature (C)") +
      xlab("Month") +
        xlim(c("5","6","7","8","9", "10", "11")) +
          labs(color="Lake name") +
            theme(legend.position = "none") +
        scale_color_brewer(palette = "Dark2")
# print(NTL_LTER_temp_boxplot)


#b. TP
NTL_LTER_TP_boxplot <-
```

```
  ggplot(PeterPaul.chem.nutrients.data, aes(x=month, y=tp_ug)) +
    geom_boxplot(aes(color=lakename)) +
      ylab("TP") +
        xlab("Month") +
           xlim(c("5","6","7","8","9", "10", "11")) +
   theme(legend.position = "none") +
         scale_color_brewer(palette = "Dark2")
# print(NTL_LTER_TP_boxplot)

#c. TN
NTL_LTER_TN_boxplot <-
  ggplot(PeterPaul.chem.nutrients.data, aes(x=month, y=tn_ug)) +
    geom_boxplot(aes(color=lakename)) +
      ylab("TN") +
        xlab("Month") +
          xlim(c("5","6","7","8","9", "10", "11")) +
          scale_color_brewer(palette = "Dark2") +
            theme(legend.position = "none")
# print(NTL_LTER_TN_boxplot)


legend_NTL_LTER <- get_legend(
  NTL_LTER_temp_boxplot +
    guides(color = guide_legend(nrow = 1)) +
    theme(legend.position = "bottom")
)

plot_grid(NTL_LTER_temp_boxplot, NTL_LTER_TP_boxplot,
          NTL_LTER_TN_boxplot, legend_NTL_LTER, nrow = 4, align = 'hv',
            axis="lbt", rel_heights = c(3, 3, 3,.25))
```

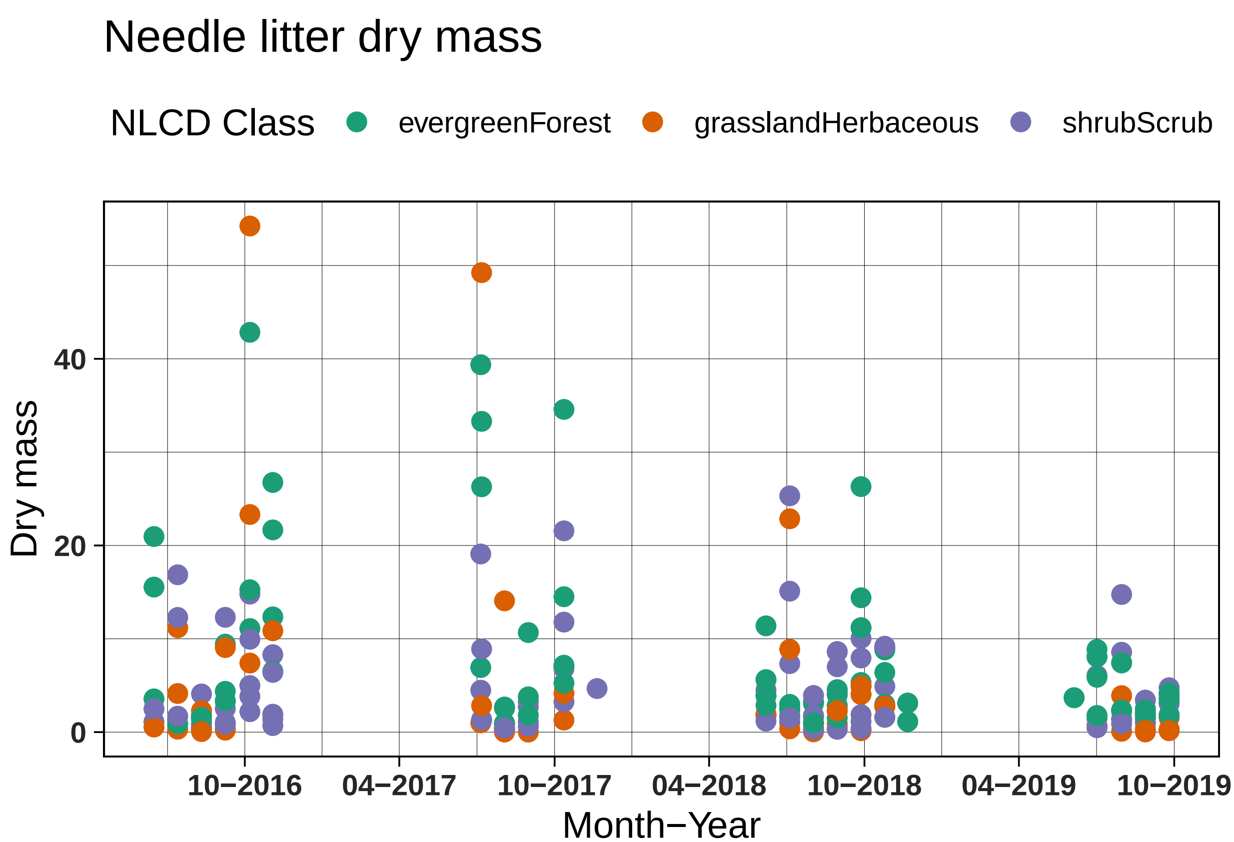


Question: What do you observe about the variables of interest over seasons and between lakes?

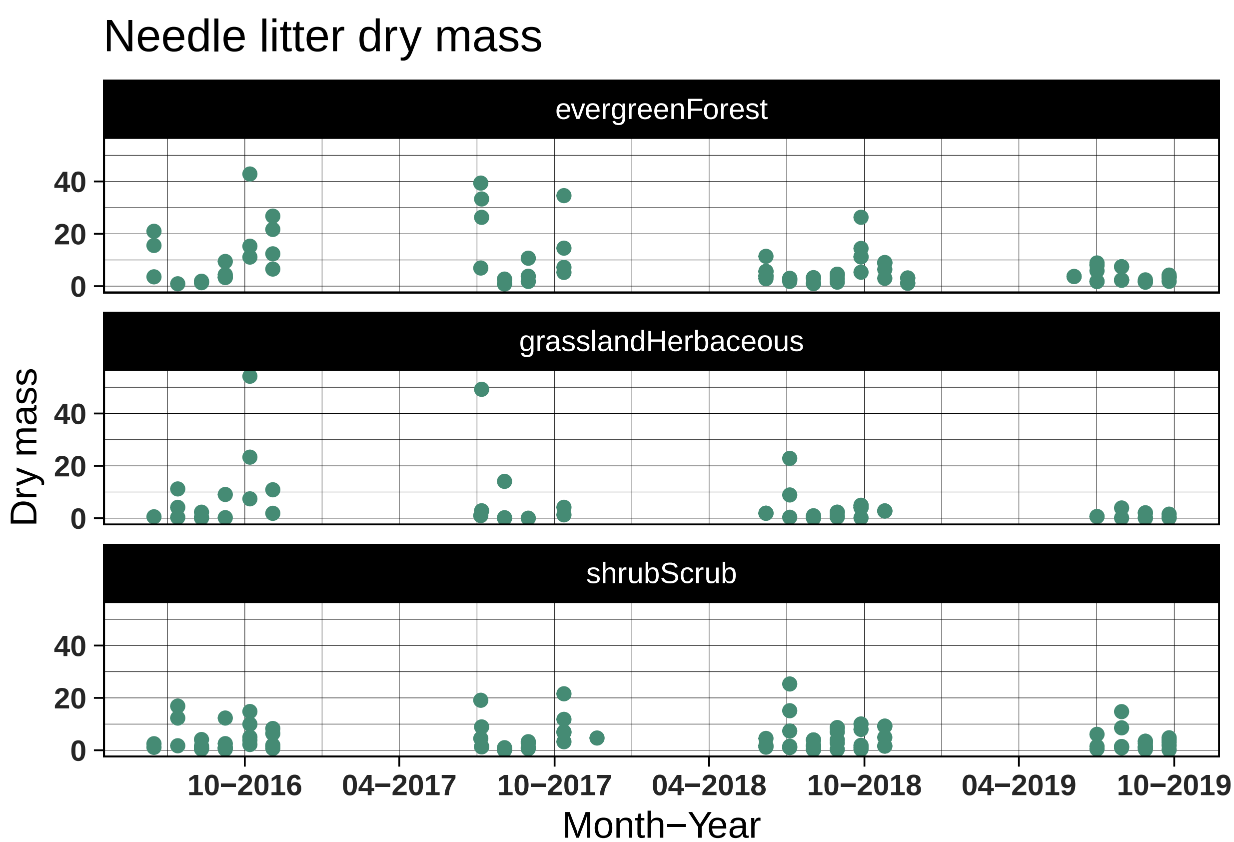
> Answer: Lake temperatures are consistent for both Paul and Peter Lake and are highest in the summer months. For both TP and TN, Peter Lake has more variablility and outliers than Paul Lake. In the summer months, Peter Lake also has higher median TP and TN than Paul Lake.

6. [Niwot Ridge] Plot a subset of the litter dataset by displaying only the "Needles" functional group. Plot the dry mass of needle litter by date and separate by NLCD class with a color aesthetic. (no need to adjust the name of each land use)
7. [Niwot Ridge] Now, plot the same plot but with NLCD classes separated into three facets rather than separated by color.

```
#6
#display.brewer.all()

NIWO.needle.subset <-
    NIWO.litter.data %>%
      filter(functionalGroup == "Needles")

NIWO.needle.plot <-
ggplot(NIWO.needle.subset, aes(x=collectDate, y=dryMass, col=nlcdClass)) +
  geom_point(size=3) +
    labs(color = "NLCD Class") +
      ggtitle("Needle litter dry mass") +
        xlab("Month-Year") +
          ylab("Dry mass") +
            scale_x_date(date_breaks="6 months", date_labels="%m-%Y") +
              scale_color_brewer(palette = "Dark2")
print(NIWO.needle.plot)
```



```
#7
NIWO.needle.facet.plot <-
ggplot(NIWO.needle.subset, aes(x=collectDate, y=dryMass)) +
  geom_point(color="aquamarine4", size=2) +
    labs(color = "NLCD Class") +
      ggtitle("Needle litter dry mass") +
        xlab("Month-Year") +
          ylab("Dry mass") +
            scale_x_date(date_breaks="6 months", date_labels="%m-%Y") +
              facet_wrap(vars(nlcdClass), nrow=3)
print(NIWO.needle.facet.plot)
```



Question: Which of these plots (6 vs. 7) do you think is more effective, and why?

> Answer: I think that either plot could be effective depending on what you are trying to convey about the dataa or analyze. If you are interested in showing how the three land types have similar dry masses, then plot 6 would be a good option becuase when overlayed is shows how the three land types have very similar ranges. However, if you want to focus on the differences between the land type dry masses, as opposed to the similarities, then plot 7 is more effective because the ranges of the dry masses overlap so much that simply color coding them is not enough, and separating them into facets will be the easiest way to visually distinguish these data.